It doesn’t usually take very long to create an FPGA design. Recently, however, a Xilinx competitor ran an ad declaring that while an FPGA can take up to a year to design, it can be cloned in only a second. Are FPGA designs really that insecure? While the ad seems absurdly hyperbolic, it is true that the bitstreams of some volatile FPGAs can be cloned. While it’s unlikely that cloning could happen in "a second," fears about the insecurity of design efforts are valid ones. To alleviate these anxieties, this white paper will show you how to substantially secure the bitstream and the overall design of FPGAs using Xilinx CoolRunner™-II CPLDs. It will not particularly address Virtex™-II and its successors because they already employ Triple DES bitstream encryption, which is considered by many to be sufficiently strong encryption to deter IP theft. It will rather focus on volatile FPGAs in general, including both Xilinx FPGAs and competitors’ FPGAs within its scope.

The Problem

Figure 1 shows a common way to download an FPGA from a standard EPROM, using a CPLD as the EPROM controller. The left hand side shows bitstream delivery directly from the EPROM into the FPGA, with the CPLD managing the addressing and delivery of control strobes to the FPGA. The right hand version has the CPLD again managing the data, but in this situation it also serializes the data and may even drive configuration through the JTAG port. The “one second” clone would be simply copying the EPROM, which in reality can take a lot more than one second. So, the task is to make it harder to “clone.” Note that the EPROM size and CPLD capacity will be chosen to provide sufficient capacity to satisfy the FPGA needs as well as other needs within the system. There may be multiple FPGAs chained for bitstream delivery, so Figure 1 is really a talking-point diagram and may be a substantially simpler configuration than many of today’s systems actually use.

First Step — External Bitstream Encryption

To thwart cloners, we recommend encrypting the EPROM bitstream with software before programming the EPROM. Then, while the CoolRunner-II loads the FPGA, it decrypts the bitstream at the same time. Cloners can’t copy the CPLD and the EPROM contents will be gibberish.

CoolRunner-II CPLDs are nonvolatile, and have many of the key features needed to support both encryption and decryption. The nonvolatility means they power up with internal bits that can be interpreted as “keys” or “passwords,” or initialization vectors, depending on your point of view. Indeed, volatility is at the heart of the security issue, and it is possible—to a degree—to convey the nonvolatility of the CoolRunner-II parts to the volatile FPGA parts, so that it will take a lot of time and money to clone the FPGA design. For conceptual reasons, this white paper will consider a fairly simple encryption/decryption method, and point you to other literature for more elaborate methods that have been shown to provide greater security strength.

The example considered here is the synchronous stream cipher, which has both strengths and weaknesses. For simplicity, we analyze a simple Linear Feedback Shift Register (LFSR), which is known to be able to calculate a pseudo-random number within its structure. Figure 2 shows a simple LFSR. This is not specifically maximal length, but shows the basic structure.
Most engineers have some exposure to LFSRs from logic design classes and possibly from communication coding theory. They are also frequently used to calculate checksums for data packets. LFSRs fall in the category of Linear Sequential Machines, which means they are constructed (typically) from D flip-flops and EX-OR gates. If initialized with a nonzero value, they will create a set of binary numbers within the flops, which will not repeat for a long clock sequence, depending on the location of the EX-OR gates and the number of D flip-flops. Identifying the feedback (aka “tap”) sites to achieve the maximal length sequence is important. Xilinx has published a number of application notes regarding LFSR behavior, referenced at the end of this document.

One of the oldest ciphers (not particularly a stream cipher) is the One Time Pad, which appeared around World War I, and was subsequently shown by Claude Shannon to be a perfect cipher, with just a few caveats. Figure 3 shows a simplistic OTP.

**Figure 2: Simple Linear Feedback Shift Register**

In Figure 3, we show the plain text (not encrypted) as a set of ones and zeroes. Then, we show the key as an array of question marks, as it is random and kept secret. When the plain text is EX-ORed with the corresponding key bits, the uncertainty of the result produces the array of question marks on the right. The requirements to be considered are the following:

1. The key length must match the message length
2. The key must be random

**Figure 3: Idea Behind a One Time Pad**
3. The key must not be re-used
Modern cryptographers have learned much, and tend to disregard the OTP as being particularly strong largely because the caveats are too tempting to violate. Long messages require long keys, which are a hassle in general. Obtaining a totally random key is also painful. Given the first two issues, there is a tendency to reuse keys. Nonetheless, for the sake of argument, let’s suppose that we can fairly easily obtain a long random key, and we will only use it one time, with a given bitstream.

So, with that as a target, our first requirement is to create a random number that matches (or exceeds) the size of the target bitstream capacity. Table 1 summarizes some bitstream capacities for some Virtex I parts, as a reference. These values will help us select LFSR lengths so that we can create pseudo random numbers that are appropriately long and eliminate the risk of introducing repetition into what will become the key for our cipher. With this set of bitstreams as targets to secure, we will need to identify a set of LFSR structures that will have enough bits, and a way for us to create the state machines and know they won’t repeat.

Table 1: Some Virtex bitstream Lengths

<table>
<thead>
<tr>
<th>Device</th>
<th># of Configuration Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>XCV50</td>
<td>559,200</td>
</tr>
<tr>
<td>XCV100</td>
<td>781,216</td>
</tr>
<tr>
<td>XCV150</td>
<td>1,040,096</td>
</tr>
<tr>
<td>XCV200</td>
<td>1,335,840</td>
</tr>
<tr>
<td>XCV300</td>
<td>1,751,808</td>
</tr>
<tr>
<td>XCV400</td>
<td>2,546,048</td>
</tr>
<tr>
<td>XCV600</td>
<td>3,607,968</td>
</tr>
<tr>
<td>XCV800</td>
<td>4,715,616</td>
</tr>
<tr>
<td>XCV1000</td>
<td>6,127,744</td>
</tr>
</tbody>
</table>

LFSRs fall in two categories, internal feedback and external feedback. Both can create pseudorandom sequences of $2^n-1$ patterns before repeating, if created with the correct primitive polynomial. Figure 4 shows the two types, and Table 2 gives the size (or degree) of the shift register and appropriate primitive polynomials.
Note that in Figure 4, we do not show initialization circuitry. Typically either asynchronous set/reset circuits provide this, or it occurs from configuration bits within the CoolRunner-II CPLD during power-on initialization.

**Table 2:** Some Primitive Polynomials and their Degree

<table>
<thead>
<tr>
<th>Degree(n)</th>
<th>Polynomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,3,4,6,7,15,22</td>
<td>$x^n + x + 1$</td>
</tr>
<tr>
<td>5,11,21,29</td>
<td>$x^n + x^2 + 1$</td>
</tr>
<tr>
<td>8,19</td>
<td>$x^n + x^6 + x^5 + x + 1$</td>
</tr>
<tr>
<td>9</td>
<td>$x^n + x^4 + 1$</td>
</tr>
<tr>
<td>10,17,20,25,28</td>
<td>$x^n + x^3 + 1$</td>
</tr>
<tr>
<td>12</td>
<td>$x^n + x^7 + x^4 + x^3 + 1$</td>
</tr>
<tr>
<td>13,24</td>
<td>$x^n + x^4 + x^3 + x + 1$</td>
</tr>
<tr>
<td>14</td>
<td>$x^n + x^{12} + x^{11} + x + 1$</td>
</tr>
<tr>
<td>16</td>
<td>$x^n + x^5 + x^3 + x^2 + 1$</td>
</tr>
</tbody>
</table>
The formula for the maximal length of the LFSR sequence given that it is derived from a primitive polynomial is simply $2^n - 1$, where “$n$” is the degree of the primitive polynomial, and we assume that it is initialized to a nonzero value. That being the case, here is how it can be used. Table 1 says that the bitstream for an XCV150 is 1,040,096 bits long. That means we need a primitive polynomial that is at least that long to create the pseudo-random number. A thousand (decimal) is a little less than 1024, which takes 10 bits to represent (0-1023). A million will take a little less than 20 bits to represent (1,048,576). That means the polynomial which has 20 bits will create a pseudo-random bitstream that is greater than 1,040,096 bits. Hence, choose $n=20$ and the corresponding polynomial would be: $X^{20} + X^3 + 1$. Figure 5 shows a corresponding LFSR. Of particular note is that it only takes a single EX-OR gate. In general, LFSRs that require minimal logic circuitry are attractive.

### Table 2: Some Primitive Polynomials and their Degree (Continued)

<table>
<thead>
<tr>
<th>$n$</th>
<th>Polynomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>$X^n + X^7 + 1$</td>
</tr>
<tr>
<td>23</td>
<td>$X^n + X^5 + 1$</td>
</tr>
<tr>
<td>26,27</td>
<td>$X^n + X^8 + X^7 + X + 1$</td>
</tr>
<tr>
<td>30</td>
<td>$X^n + X^{16} + X^{15} + X + 1$</td>
</tr>
</tbody>
</table>
The circuit in Figure 5 would take up the flip-flops for 20 macrocells, and would thus fit easily into the smallest CoolRunner-II CPLD. In CoolRunner-II CPLDs, each macrocell is comprised of a flip-flop and a set of logic driving it that can create a Sum of Products logic structure. Also contained within the macrocell is an EX-OR gate. The architecture is clustered into 16-macrocell function blocks, so this function would partially consume 1.25 FB. That leaves much logic available to form other functions, as needed.

Figure 5:  Internal Feedback LFSR for Primitive Polynomial $X^{20} + X^3 + 1$
(clock and initialization signals not shown)

Attacking the Bitstream Encryption

Now that we have constructed a proposal for using an LFSR to decrypt the bitstream, let’s consider how it may be attacked by a “cloner.” Just to reiterate, we are assuming that an FPGA design has been created and a corresponding bitstream produced. The bitstream was subsequently run through a software routine (typically), and has EX-ORed successive bits with an internal model of an appropriate LFSR. This resulted in an encrypted bitstream that is loaded into an EPROM.

It would be the developer’s job to ensure that the bitstream calculation was done correctly, as was the creation of the appropriate file formatted for downloading (checksum calculations, etc.) When the EPROM contents are accessed by the CPLD for
subsequent delivery into the FPGA, they are EX-ORed with a hardware version of the LFSR, which produces clear bits for delivery to the FPGA.

How might this be attacked? There are several ways, but there are two that are most typical. First, an attacker could create a circuit capable of capturing the clear bits coming out of the CPLD. Figure 6 shows a simplified circuit for doing that.

![Diagram of pipeline installation](image)

Figure 6: Rough Circuit to Capture bitstream "On the Fly"

In Figure 6, the CPLD with LFSR delivers its clear bits (serial or parallel) to the FPGA, while CPLD #2 contains a circuit that copies the bits as they pass by into a shadow SRAM. Details for CPLD #2 and the SRAM depth will depend on the bitstream being captured, but it will still require some serious training for the attacker to capture all the bits, in the right format, protocol, etc. Being off by a single shift out of a million plus bits won’t work. Any noise, clock jitter, or any one of a number of other practical issues makes this a lot tougher than it looks. Then, the data would need to be formatted and driven into the FPGA to assure it works. It has its merit, but definitely takes more than a “second to clone.”

The other general way to attack this would be to recognize in general the way the CPLD is working and to subvert it. If an EPROM with all zeroes is inserted into the EPROM position (remove chip and insert blank), then an interesting fact about the EX-OR function is discovered: A EX-OR 0 = A. That’s right. A blank EPROM would divulge the pseudo-random number being created by the CPLD’s LFSR. Simple LFSRs have had too much history! Being linear state machines, they have an impulse response. If you know that, you can convolve (logical convolution) the impulse
response with an input stream and predict the results. This has led to ways to identify the primitive polynomial very systematically, which means it can be cracked by a set of experiments.

There are other kinds of attacks. There are correlation attacks, linear syndrome attacks and a wide assortment of others. Luckily, there are also a bunch of circuits that have been developed for building stream ciphers that are much harder to crack. Those are described in an appendix, later.

Second Step — Simple Steganography

Steganography is basically hiding one message inside another. In our situation, we have identified a weakness of our LFSR—the blank EPROM attack. Although we will be recommending a more elaborate stream cipher solution, it won’t hurt to add in protection against the blank EPROM.

Again, the idea is simple. The CPLD won’t use up all its logic doing the stream cipher, so it makes sense to take out some insurance and add more circuitry to offer a change of response if being attacked. We could simply scan through the EPROM and if all addresses have zeroes, inhibit the delivery of the bitstream through the stream cipher. An attacker would then observe that we ran through all the addresses and didn’t deliver anything to the output. That would work. Alternately, we could scan the EPROM and deliver a different but also pseudo-random pattern (say, from a different internal LFSR polynomial) to the FPGA. Doing so might even destroy the FPGA. This may or may not be a good thing.

A third approach would be to scan the EPROM and look for a specific set of bits at particular locations. This means that we would read the EPROM, and when certain addresses are created, we would compare what we found in the EPROM to internal copies of the same bits within the CPLD. If all of the targeted address contents matched the internal CPLD copies, the next round of addressing would be collecting the real encrypted EPROM data, and delivering it to the FPGA. The particular addresses we choose would constitute our “steganographic” authentication message: "Yes, this is a correct EPROM, so go ahead and decrypt it and load the FPGA."

At this point, we have done something that is suspicious: we have introduced a complete scan of the EPROM before we start delivering the decrypted data. We could add confusion here by using an LFSR to pseudo-randomly produce EPROM addresses, and again look for contents of specific address locations. If we did the initial scan pseudo-randomly, then did a linear addressing when we read out the encrypted data, an attacker could identify when we switch from a scan phase to a decrypt phase. So, if we chose to do the scan for correct EPROM data, then subsequently did a pseudo-random access of the EPROM for the decrypting, it would be harder for an attacker to discover what is going on. It also means the encrypted bitstream would have to be loaded into the EPROM file in an interesting order that might take a while to get right. But, it has merit for thwarting attackers.

Attacking the Steganography

It is difficult to discover the steganography, but not impossible. As described above, there are two distinct reading phases. The first, where the CPLD simply looks through the EPROM for certain items, must be done completely. If any addresses are skipped, then that narrows the search for what is being looked at. Make sure the CPLD examines every address within the EPROM, and doesn’t take any action differently until everything has been looked at. Short cycling the search can result in ultimately divulging the decision points, which can be a weakness with regard to Differential Power Analysis (DPA), and Tempest attacks.
DPA means that an attacker could inspect the power-supply pins of the CPLD, observe its behavior when scanning a blank EPROM, scan its behavior with a valid EPROM, and compare addresses and data when there are differences. If the attacker can discover which addresses are being scanned for, he can then do a brute force attack on those addresses to find which pass the steganography test. If all addresses are looked at before a conclusion is reached, this makes the decision for multiple data items occur only once, and always at the end. It is best, however, even given that, to still make sure the key inspected data is scattered through the middle of the EPROM address space.

There are also other subtle techniques along these lines, such as using neural nets. Tempest is basically doing a similar inspection of the electromagnetic radiation coming off the CPLD. All of these attacks have merit. Luckily, they are also very hard to accomplish and sort out. See References, page 11 at the end for additional information.

The Third Step—Repartitioning

As commented earlier, frequently CPLDs are on the board doing different tasks than loading the FPGA. They are programmable logic devices with their own characteristics and abilities, among which is relatively fast pin to pin speeds and high FMAX operation. That being the case, their nonvolatility can be used in collaboration with the volatile requirements of SRAM based FPGAs. Design partitioning is basically separating aspects of a given design and distributing it among the various logic components within the system being designed. The recommendation here is simple. Inspect the FPGA design and identify a section of it that will be critical for the overall operation, but can actually be placed in the CPLD instead of in the FPGA. This may mean redefining the FPGA functionality, but here is a hint: FPGAs are extremely good at high speed, regular arithmetic operations, and those aspects frequently are best held within the FPGA. However, the control function for those aspects might be good operations to move into the CPLD.

Conclusion

We have discussed general techniques to make it hard for cloners to obtain FPGA bitstreams. We discussed combining three techniques—cryptography, steganography and partitioning—to raise the bar for cloners. The circuitry for each is fairly small, especially in light of the additional security they will bring to your system.

It would be possible to gain extra security by adopting just cryptography, or just partitioning (although the use of all of the three techniques together will keep your system best protected). If a key element of the FPGA is within the CPLD, and its security bits are set, then it will require reverse engineering the CPLD to get the rest of the design—assuming, of course, that the function contained in the CPLD is not so obvious that it can be deduced by observation. CoolRunner-II CPLDs add substantially to the security picture, when properly used! Be sure and read more about the CoolRunner-II security in Xilinx White Paper WP170.

Appendix: Additional Steps for the Insecure and Paranoid

As mentioned earlier, LFSRs are considered to be cryptographically insecure when directly building stream ciphers. The references include several options which go beyond simple LFSR solutions and appear to have merit as viable stream ciphers. In particular, Reference 1 summarizes more than a dozen that combine LFSRs in interesting ways to increase the resistance to attack. Among them are:

- Bilateral Stop and Go generators
- Dynamic random sequence generator
• Gollman cascade
• Shrinking generator
• Self shrinking generator

Typically, these ciphers combine several LFSRs by sequentially stacking them, by combining several with multiplexers, by logically combining multiple LFSRs to direct the clocking of others, etc.

Going beyond cryptography, steganography and partitioning, there are other actions that can be taken. Using chip scale packages with balls that hug the board surface makes it difficult to probe signals. Connecting PCB traces on internal traces, without surfacing except right at the pin sites, also makes it difficult. Employing conformal chip coatings is still another way of keeping signals secure, and it is certain that even more techniques exist.

References

Stream Ciphers


LFSRs.


Stream Ciphers Attacks


27. "On the Applicability of Distinguishing Attacks Against Stream Ciphers," Greg Rose, Philip Hawkes, Third Nessie Workshop


### Steganography


### Randomness


### Further Reading

**Application Notes**


CoolRunner-II Data Sheets


CoolRunner-II White Papers


Revision History

The following table shows the revision history for this document.

<table>
<thead>
<tr>
<th>Date</th>
<th>Version</th>
<th>Revision</th>
</tr>
</thead>
<tbody>
<tr>
<td>06/30/03</td>
<td>1.0</td>
<td>Initial Xilinx release.</td>
</tr>
</tbody>
</table>